

An Exposition of Götze's Estimation of the Rate of Convergence in the Multivariate Central Limit Theorem

Rabi Bhattacharya *

and Susan Holmes †

Department of Mathematics

Math 603

University of Arizona

e-mail: rabi@math.arizona.edu

Department of Statistics

Sequoia Hall

Stanford University

CA 94305

e-mail: susan@stat.stanford.edu

url: <http://www-stat.stanford.edu/~susan>

Abstract: We provide an explanation of the main ideas underlying Götze's main result in [9] using Stein's method. We also provide detailed derivations of various intermediate estimates. Curiously, we are led to a different dimensional dependence of the constant than that given in [9].

We would like to dedicate this to Charles Stein on the occasion of his 90th birthday.

AMS 2000 subject classifications: Primary 60F05, 60B10; secondary 62H05.

Keywords and phrases: Multivariate CLT, Stein's Method.

*Supported by NSF DMS-0806011

†Supported by NSF DMS-02 41246 and NIH R01GM086884-02

1. Introduction

In his article Götze[9] used Stein's method to provide an ingenious derivation of the Berry-Esseen type bound for the class of Borel convex subsets of \mathbb{R}^k in the context of the classical multivariate central limit theorem (CLT). This approach has proved fruitful in deriving error bounds for the CLT under certain structures of dependence as well (see Rinott and Rotar [11]). Our view and elaboration of Götze's proof resulted from a collaboration between the authors and were first presented in a seminar at Stanford given by the first author in the summer of 2000. The authors wish to thank Persi Diaconis for pointing out the need for a more readable account of Götze's result than given in his original work.

After an explanation of the general method in Section 1, detailed derivations of various estimates are given in Sections 2-4 in terms that would be reasonably familiar to probabilists. Except for the smoothing inequality in Section 4, which is fairly standard, complete proofs are given.

Recently Raic[10] has followed essentially the same route as Götze, but in greater detail, in deriving Götze's bound. It may be pointed out that we were unable to verify the dimensional dependence $O(k)$ in [9],[10]. Our derivation provides the higher order dependence of the error rate on k , namely $O(k^{\frac{5}{2}})$. This rate can be reduced to $O(k^{\frac{3}{2}})$ using an inequality of Ball [1]. The best order of dependence known, namely, $O(k^{\frac{1}{4}})$ is given by Bentkus[3], using a different method, which would be difficult to extend to dependent cases.

As a matter of notation, the constants c , with or without subscripts are absolute constants. The k -dimensional standard Normal distribution is denoted by $\mathcal{N}(0, \mathbb{I}_k)$ as well as Φ , with density ϕ .

1.1. The Generator of the ergodic Markov process as a Stein operator.

Suppose Q and Q_0 are two probability measures on a measurable space (S, \mathcal{S}) and h is integrable (with regards to Q and Q_0). Consider the problem of estimating

$$(1.1) \quad Eh - E_0h \equiv \int h dQ - \int h dQ_0.$$

A basic idea of Stein[12] (developed in some examples in [7] and [8]) is

- (i) to find an invertible map L which maps "nice" functions on S into the kernel or null space of E_0 ,
- (ii) to find a perturbation of L , say L_α , which maps "nice" functions on S into the kernel or null space of E ,
- (iii) to estimate 1.1 using the identity

$$(1.2) \quad Eh - E_0h = ELg_0 = E(Lg_0 - L_\alpha g_\alpha)$$

where

$$g_0 \equiv L^{-1}(h - E_0h), \quad g_\alpha \equiv L_\alpha^{-1}(h - Eh).$$

One way to find L is to consider an ergodic Markov process $\{X_t : t \geq 0\}$ on S which has Q_0 as its invariant distribution, and let L be its generator:

$$(1.3) \quad Lg = \lim_{t \downarrow 0} \frac{T_t g - g}{t}, \quad g \in \mathcal{D}_L$$

where the limit is in $L^2(S, Q_0)$, and

$$(T_t g)(x) = E[g(X_t) | X_0 = x],$$

or in terms of the transitions probability $p(t; x, dy)$ of the Markov process $\{X_t : t \geq 0\}$,

$$(1.4) \quad (T_t g)(x) = \int_S g(y) p(t; x, dy) \quad (x \in S, t > 0).$$

Also \mathcal{D}_L is the set of g for which the limit in (1.3) exists. By the Markov (or, semigroup) property, $T_{t+s} = T_t T_s = T_s T_t$, so that

$$(1.5) \quad \frac{d}{dt} T_t g = \lim_{s \downarrow 0} \frac{T_{t+s} g - T_t g}{s} = \lim_{s \downarrow 0} \frac{T_t (T_s g - g)}{s} = T_t Lg.$$

Since $T_t T_s = T_s T_t$, T_t and L commute so that

$$(1.6) \quad \frac{d}{dt} T_t g = L T_t g.$$

Note that invariance of Q_0 means $E T_t g(X_0) = E g(X_0) = \int g dQ_0$, if the distribution of X_0 is Q_0 . This implies that, for every $g \in \mathcal{D}_L$, $E Lg(X_0) = 0$, or

$$\int_S Lg(x) dQ_0(x) = 0, \left[E Lg(X_0) = E \left(\lim_{t \downarrow 0} \frac{T_t g(X_0) - g(X_0)}{t} \right) = \lim_{t \downarrow 0} \frac{E T_t g(X_0) - E g(X_0)}{t} \right]$$

That is, L maps \mathcal{D}_L into the set 1^\perp of mean zero functions in $L^2(S, Q_0)$. It is known that the range of L is dense in 1^\perp and if L has a spectral gap, then the range of L is all of 1^\perp . In the latter case L^{-1} is well defined on 1^\perp (kernel of Q_0) and is bounded on it ([4]).

Since T_t converges to the identity operator as $t \downarrow 0$ one may also use T_t for small $t > 0$ to smooth the target function $\tilde{h} = h - \int h dQ_0$. For the case of a diffusion $\{X_t : t \geq 0\}$, L is a differential operator and even non smooth functions such as $\tilde{h} = \mathbf{1}_B - Q_0(B)(h = \mathbf{1}_B)$ are immediately made smooth by applying T_t . One may then use the approximation to \tilde{h} given by

$$(1.7) \quad T_t \tilde{h} = L(L^{-1} T_t \tilde{h}) = L \psi_t, \text{ with } \psi_t = L^{-1} T_t \tilde{h},$$

and then estimate the error of this approximation by a “smoothing inequality”, especially if $T_t \tilde{h}$ may be represented as a perturbation by convolution. For several perspectives and applications of Stein's method see [2], [7],[8],[11].

1(b) The Ornstein-Uhlenbeck Process and its Gausssian invariant Distribution

The Ornstein-Uhlenbeck (O-U) process is governed by the Langevin equation (see, e.g. [6, pp. 476, 597, 598])

$$(1.8) \quad dX_t = -X_t dt + \sqrt{2} dB_t$$

where $\{B_t : t \geq 0\}$ is a k -dimensional standard Brownian motion. Its transition density is

$$(1.9) \quad p(t; x, y) = \prod_{i=1}^k [2\pi(1 - e^{-2t})]^{-\frac{1}{2}} \exp\left\{-\frac{(y_i - e^{-t}x_i)^2}{2(1 - e^{-2t})}\right\} \quad x = (x_1, \dots, x_k), y = (y_1, \dots, y_k).$$

This is the density of a Gaussian (Normal) distribution with mean vector $e^{-t}x$ and dispersion matrix $(1 - e^{-2t})\mathbb{I}_k$ where \mathbb{I}_k is the $k \times k$ identity matrix. One can check (e.g., by direct differentiation) that the Kolmogorov backward equation holds:

$$(1.10) \quad \frac{\partial p(t; x, y)}{\partial t} = \sum_{i=1}^k \frac{\partial^2 p(t; x, y)}{\partial x_i^2} - \sum_{i=1}^k x_i \frac{\partial p(t; x, y)}{\partial x_i} = \Delta p - x \bullet \nabla p = Lp, \text{ with } L \equiv \Delta - x \bullet \nabla$$

where Δ is the Laplacian and $\nabla = \text{grad}$. Integrating both sides w.r.t. $h(y)dy$ we see that $T_t h(x) = \int h(y)p(t; x, y)dy$ satisfies

$$(1.11) \quad \frac{\partial}{\partial t} T_t h(x) = \Delta T_t h(x) - x \bullet \nabla T_t h(x) = L T_t h(x), \quad \forall h \in L^2(\mathbb{R}^k, \Phi).$$

Now on the space $L^2(\mathbb{R}^k, \Phi)$ (where $\Phi = N(0, \mathbb{I}_k)$ is the k -dimensional standard Normal), L is self adjoint and has a spectral gap, with the eigenvalue 0 corresponding to the invariant distribution Φ (or the constant function $\mathbf{1}$ on $L^2(\mathbb{R}^k, \Phi)$). This may be deduced from the fact that the Normal density $p(t; x, y)$ (with mean vector $e^{-t}x$ and dispersion matrix $(1 - e^{-2t})\mathbb{I}_k$) converges to the standard Normal density $\phi(y)$ exponentially fast as $t \rightarrow \infty$, for every initial state x . Else, one can compute the set of eigenvalues of L , namely $\{0, -1, -2, \dots\}$ — with eigenfunctions expressed in terms of Hermite polynomials [6, page 487]. In particular, L^{-1} is a bounded operator on 1^\perp and is given by

$$(1.12) \quad L^{-1} \tilde{h} = - \int_0^\infty T_s \tilde{h}(x) ds, \quad \forall \tilde{h} = h - \int h d\Phi \in L^2(\mathbb{R}^k, \Phi).$$

To check this, note that by (1.11)

$$(1.13) \quad \tilde{h} = - \int_0^\infty \frac{\partial}{\partial s} T_s \tilde{h}(x) ds = - \int_0^\infty L T_s \tilde{h}(x) ds = L \left(- \int_0^\infty T_s \tilde{h}(x) ds \right).$$

For our purposes $h = \mathbf{1}_C$: the indicator function of a Borel convex subset C of \mathbb{R}^k .

A smooth approximation of \tilde{h} is $T_t \tilde{h}$ for small $t > 0$ (since $T_t \tilde{h}$ is infinitely differentiable). Also, by (1.12)

$$(1.14) \quad \begin{aligned} \psi_t(x) \equiv L^{-1} T_t \tilde{h}(x) &= - \int_0^\infty T_s T_t \tilde{h}(x) ds = - \int_0^\infty T_{s+t} \tilde{h}(x) ds = - \int_t^\infty T_s \tilde{h}(x) ds \\ &= - \int_t^\infty \left\{ \int_{\mathbb{R}^k} \tilde{h}(e^{-s}x + \sqrt{1 - e^{-2s}}z) \phi(z) dz \right\} ds \end{aligned}$$

where ϕ is the k -dimensional standard Normal density. We have expressed $T_s \tilde{h}(x) \equiv E[\tilde{h}(X_s) | X_0 = x]$ in (1.14) as

$$(1.15) \quad E[\tilde{h}(X_s) | X_0 = x] = E\tilde{h}(e^{-s}x + \sqrt{1 - e^{-2s}}Z),$$

where Z is a standard Normal $N(0, \mathbb{I}_k)$. For X_s has the same distribution as $e^{-s}x + \sqrt{1 - e^{-2s}}Z$. Now note that using (1.14), one may write

$$(1.16) \quad T_t \tilde{h}(x) = L(L^{-1} T_t \tilde{h}(x)) = \Delta(L^{-1} T_t \tilde{h}(x)) - x \bullet \nabla(L^{-1} T_t \tilde{h}(x)) = \Delta \psi_t(x) - x \bullet \nabla \psi_t(x).$$

For the problem at hand (see 1.1) $Q_0 = \Phi$ and $Q = Q_{(n)}$ is the distribution of $S_n = \frac{1}{\sqrt{n}}(Y_1 + Y_2 + \cdots + Y_n) = (X_1 + X_2 + \cdots + X_n)/\sqrt{n}$, where Y_j 's are i.i.d. mean-zero with covariance matrix \mathbb{I}_k and finite absolute third moment

$$\rho_3 = E\|Y_1\|^3 = E\left(\sum_{i=1}^k (Y_1^{(i)})^2\right)^{\frac{3}{2}}.$$

We want to estimate

$$(1.17) \quad E\tilde{h}(S_n) = Eh(S_n) - \int h d\Phi$$

for $h = \mathbf{1}_C$, $C \in \mathcal{C}$ —the class of all Borel convex sets in \mathbb{R}^k .

For this we first estimate (see (1.16)), for small $t > 0$,

$$(1.18) \quad ET_t\tilde{h}(S_n) = E[\Delta\psi_t(S_n) - S_n \bullet \nabla\psi_t(S_n)]$$

This is done in Section 3. The next step is to estimate, for small $t > 0$,

$$(1.19) \quad ET_t\tilde{h}(S_n) - E\tilde{h}(S_n)$$

which is carried out in Section 4. Combining the estimates of (1.18) and (1.19), and with a suitable choice of $t > 0$, one arrives at the desired estimation of (1.17).

We will write

$$(1.20) \quad \delta_n = \sup_{\{h=\mathbf{1}_C: C \in \mathcal{C}\}} \left| \int h dQ_{(n)} - \int h d\Phi \right|.$$

2. Derivatives of $\psi_t \equiv L^{-1}T_t\tilde{h}$

Before we engage in the estimation of (1.18) and (1.19), it is useful to compute certain derivatives of ψ_t .

$$\text{Let } D_i = \frac{\partial}{\partial x_i}, \quad D_{ii'} = \frac{\partial^2}{\partial x_i \partial x_{i'}}, \quad D_{ii'i''} = \frac{\partial^3}{\partial x_i \partial x_{i'} \partial x_{i''}}, \text{ etc. .}$$

Then, using (1.14),

$$\begin{aligned} D_i\psi_t(x) &= - \int_t^\infty \left[\int_{\mathbb{R}^k} \tilde{h}(y) (2\pi(1-e^{-2s}))^{-\frac{k}{2}} \frac{(y_i - e^{-s}x_i)}{1-e^{-2s}} e^{-s} \cdot \exp\left\{-\frac{\|y - e^{-s}x\|^2}{2(1-e^{-2s})}\right\} dy \right] ds \\ &= - \int_t^\infty \left[\int_{\mathbb{R}^k} \tilde{h}(y) (2\pi(1-e^{-2s}))^{-\frac{k}{2}} \frac{e^{-s}}{\sqrt{1-e^{-2s}}} \cdot \frac{(y_i - e^{-s}x_i)}{\sqrt{1-e^{-2s}}} \cdot \exp\left\{-\frac{\|y - e^{-s}x\|^2}{2(1-e^{-2s})}\right\} dy \right] ds \\ (2.1) \quad &= - \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right) \left[\int_{\mathbb{R}^k} \tilde{h}(e^{-s}x + \sqrt{1-e^{-2s}}z) z_i \phi(z) dz \right] ds, \quad z_i \phi(z) = -\frac{\partial}{\partial z_i} \phi(z) = -D_i \phi(z) \end{aligned}$$

using the change of variables

$$z = \frac{y - e^{-s}x}{\sqrt{1-e^{-2s}}}.$$

In the same manner, one has, using $D_i, D_{ii'}$, etc for derivatives $\frac{\partial}{\partial z_i}, \frac{\partial^2}{\partial z_i \partial z_{i'}}$, etc,

$$(2.2) \quad \begin{aligned} D_{ii'} \psi_t(x) &= - \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^2 \left[\int_{\mathbb{R}^k} \tilde{h}(e^{-s}x + \sqrt{1-e^{-2s}}z) \cdot D_{ii'} \phi(z) dz \right] ds, \\ D_{ii'i''} \psi_t(x) &= - \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 \left[\int_{\mathbb{R}^k} \tilde{h}(e^{-s}x + \sqrt{1-e^{-2s}}z) \cdot (-D_{ii'i''} \phi(z)) dz \right] ds. \end{aligned}$$

The following estimate is used in the next section:

$$(2.3) \quad \sup_{u \in \mathbb{R}^k} \left| \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \tilde{h} \left(\sqrt{\frac{n-1}{n}} e^{-s} x + e^{-s} u + \sqrt{1-e^{-2s}} z \right) \phi(x) D_{ii'i''} \phi(z) dx dz \right| \leq c_0 k e^{2s} (1 - e^{-2s}).$$

To prove this, write $a = \sqrt{\frac{n}{n-1}} e^s \sqrt{1-e^{-2s}}$ and change variables $x \rightarrow y = x + az$. Then

$$(2.4) \quad \phi(x) = \phi(y - az) = \phi(y) - az \cdot \nabla \phi(y) + a^2 \sum_{r,r'=1}^k z_r z_{r'} \int_0^1 (1-v) D_{rr'} \phi(y - vaz) dv,$$

so that

$$\tilde{h} \left(\sqrt{\frac{n-1}{n}} e^{-s} x + e^{-s} u + \sqrt{1-e^{-2s}} z \right) = \tilde{h} \left(\sqrt{\frac{n-1}{n}} e^{-s} y + e^{-s} u \right),$$

and the double integral in (2.3) becomes

$$(2.5) \quad \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \tilde{h} \left(\sqrt{\frac{n-1}{n}} e^{-s} y + e^{-s} u \right) \left[\phi(y) - az \cdot \nabla \phi(y) + a^2 \sum_{r,r'=1}^k z_r z_{r'} \int_0^1 (1-v) D_{rr'} \phi(y - vaz) dv \right] \bullet D_{ii'i''} \phi(z) dz dy$$

Note that the integrals of $D_{ii'i''} \phi(z)$ and $z_{i_0} D_{ii'i''} \phi(z)$ vanish for i, i', i'', i_0 , so that

$$(2.6) \quad \int_{\mathbb{R}^k} \tilde{h} \left(\sqrt{\frac{n-1}{n}} e^{-s} y + e^{-s} u \right) (\phi(y) - az \cdot \nabla \phi(y)) D_{ii'i''} \phi(z) dz = 0$$

The magnitude of the last term on the right in (2.4) is

$$(2.7) \quad \begin{aligned} & \left| a^2 \int_0^1 (1-v) \left[\sum_{r,r'=1}^k z_r z_{r'} (y - vaz)_r (y - vaz)_{r'} - \sum_{r=1}^k z_r^2 \right] \phi(y - vaz) dv \right| \\ & \leq a^2 \int_0^1 (1-v) \left[\sum_{r,r'=1}^k z_r z_{r'} (y - vaz)_r (y - vaz)_{r'} + \sum_{r=1}^k z_r^2 \right] \phi(y - vaz) dv, \end{aligned}$$

since the sum $\sum_{r,r'}$ above is nonnegative. Bounding $|\tilde{h}|$ by 1, it follows from (2.5)-(2.7) that the left side of (2.3) is no more than

$$(2.8) \quad \begin{aligned} & a^2 \int_0^1 (1-v) \left\{ \int_{\mathbb{R}^k} \left[\sum_{r \neq r'}^k z_r z_{r'} \int_{\mathbb{R}^k} (y - vaz)_r (y - vaz)_{r'} \phi(y - vaz) dy + \sum_{r=1}^k z_r^2 \int_{\mathbb{R}^k} \{(y - vaz)_r^2 + 1\} \phi(y - vaz) dy \right] |D_{ii'i''} \phi(z)| dz \right\} dv \\ & = a^2 \int_0^1 (1-v) \left\{ \int_{\mathbb{R}^k} 2 \sum_{r=1}^k z_r^2 |D_{ii'i''} \phi(z)| dz \right\} dv, \end{aligned}$$

from which (2.3) follows.

3. Estimation of $T_t \tilde{h}(S_n)$

By (1.16),

$$(3.1) \quad T_t \tilde{h}(S_n) = L(L^{-1}T_t \tilde{h})(S_n) = L\psi_t(S_n) = \Delta\psi_t(S_n) - S_n \bullet \nabla\psi_t(S_n)$$

Consider the Taylor expansions

$$(3.2) \quad \begin{aligned} \Delta\psi_t(S_n) &\equiv \sum_{i=1}^k D_{ii}\psi_t(S_n) = \sum_{i=1}^k D_{ii}\psi_t(S_n - X_1) + \sum_{i,i'=1}^k \int_0^1 X_1^{(i')} D_{iii'}\psi_t(S_n - X_1 + vX_1)dv, \\ S_n \cdot \nabla\psi_t(S_n) &= \sum_{j=1}^n X_j \cdot \nabla\psi_t(S_n) = \sum_{j=1}^n \sum_{i=1}^k X_j^{(i)} \cdot D_i\psi_t(S_n) \\ &= \sum_{j=1}^n \left[\sum_{i=1}^k X_j^{(i)} \cdot D_i\psi_t(S_n - X_j) + \sum_{i,i'=1}^k X_j^{(i)} X_j^{(i')} D_{ii'}\psi_t(S_n - X_j) + \right. \\ &\quad \left. + \sum_{i,i',i''=1}^k X_j^{(i)} X_j^{(i')} X_j^{(i'')} \int_0^1 (1-v) D_{ii'i''}\psi_t(S_n - X_j + vX_j)dv \right] \end{aligned}$$

Recalling that $X_j = \frac{Y_j}{\sqrt{n}}$, $EY_j = 0$, $EX_j^{(i)} X_j^{(i')} = \frac{1}{n} EY_j^{(i)} Y_j^{(i')} = \frac{1}{n} \delta_{ii'}$ and X_j and $S_n - X_j$ are independent,

$$(3.3) \quad E\Delta\psi_t(S_n) = E \left[\sum_{i=1}^k D_{ii}\psi_t(S_n - X_1) \right] + E \left[\sum_{i,i'=1}^k \frac{Y_1^{(i)}}{\sqrt{n}} \int_0^1 D_{iii'}\psi_t(S_n - X_1 + vX_1)dv \right],$$

$$(3.4) \quad ES_n \cdot \nabla\psi_t(S_n) = E \left[\sum_{i=1}^k D_{ii}\psi_t(S_n - X_1) \right] + \frac{1}{\sqrt{n}} \sum_{i,i',i''=1}^k E \left[Y_1^{(i)} Y_1^{(i')} Y_1^{(i'')} \int_0^1 (1-v) D_{ii'i''}\psi_t(S_n - X_1 + vX_1)dv \right].$$

Hence

$$(3.5) \quad ET_t \tilde{h}(S_n) = E \left[\sum_{i,i'=1}^k \frac{Y_1^{(i)}}{\sqrt{n}} \int_0^1 D_{iii'}\psi_t(S_n - X_1 + vX_1)dv - \frac{1}{\sqrt{n}} \sum_{i,i',i''=1}^k Y_1^{(i)} Y_1^{(i')} Y_1^{(i'')} \int_0^1 (1-v) D_{ii'i''}\psi_t(S_n - X_1 + vX_1)dv \right].$$

One may then write

$$(3.6) \quad ET_t \tilde{h}(S_n) = E[E(\bullet \bullet \bullet | Y_1)]$$

where $\bullet \bullet \bullet$ is the quantity within square brackets in (3.5), i.e.,

$$(3.7) \quad \begin{aligned} E[T_t \tilde{h}(S_n) | Y_1] &= \frac{1}{\sqrt{n}} \sum_{i,i'=1}^k Y_1^{(i')} \int_0^1 E[D_{iii'}\psi_t(S_n - X_1 + vX_1) | Y_1] dv \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i,i',i''=1}^k Y_1^{(i)} Y_1^{(i')} Y_1^{(i'')} \int_0^1 (1-v) E[D_{ii'i''}\psi_t(S_n - X_1 + vX_1) | Y_1] dv \end{aligned}$$

The first term on the right side in (3.7) equals

$$\begin{aligned}
 & \frac{1}{\sqrt{n}} \sum_{i,i'=1}^k Y_1^{(i')} \left(- \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 \int_0^1 \left\{ E \left[\int_{\mathbb{R}^k} \tilde{h}(e^{-s}(S_n - X_1) + e^{-s}vX_1 + \sqrt{1-e^{-2s}}z) \cdot (-D_{iii'}\phi(z))dz | Y_1 \right] \right\} dv \right) ds \\
 & (3.8) \\
 & = \frac{1}{\sqrt{n}} \sum_{i,i'=1}^k Y_1^{(i')} \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 \left(\int_0^1 \left\{ \int_{\mathbb{R}^k} \left[\int_{\mathbb{R}^k} \tilde{h}(e^{-s}\sqrt{\frac{n-1}{n}}x + e^{-s}vX_1 + \sqrt{1-e^{-2s}}z) dQ_{(n-1)}(x) \right] D_{iii'}\phi(z) dz \right\} dv \right) ds,
 \end{aligned}$$

noting that the distribution of $S_n - X_1 = \sqrt{\frac{n-1}{n}}(\frac{Y_2+Y_3+\dots+Y_n}{\sqrt{n-1}})$ is that of $\sqrt{\frac{n-1}{n}}V$, where V has distribution $Q_{(n-1)}$.

Therefore, (3.8) is equal to

$$\begin{aligned}
 & (3.9) \\
 & \frac{1}{\sqrt{n}} \sum_{i,i'=1}^k Y_1^{(i')} \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 \int_0^1 \left\{ \int_{\mathbb{R}^k} \left[\int_{\mathbb{R}^k} \tilde{h}(e^{-s}\sqrt{\frac{n-1}{n}}x + e^{-s}vX_1 + \sqrt{1-e^{-2s}}z) (d(Q_{(n-1)}(x) - \Phi(x)) + d\Phi(x)) \right] \right. \\
 & \left. \cdot D_{iii'}\phi(z) dz \right\} dv ds
 \end{aligned}$$

Since the class of functions $h = \mathbf{1}_C$, where C ranges over all Borel convex subsets of \mathbb{R}^k , is invariant under translation, and bC is convex if C is convex ($bC = \{bx : x \in C\}, \forall b > 0$),

$$(3.10) \quad \left| \int_{\mathbb{R}^k} \tilde{h}(e^{-s}\sqrt{\frac{n-1}{n}}x + e^{-s}vX_1 + \sqrt{1-e^{-2s}}z) (d(Q_{(n-1)}(x) - \Phi(x))) \right| \leq \delta_{n-1}.$$

Similarly, the second term on the right in (3.7) equals

$$\begin{aligned}
 & (3.11) \quad - \frac{1}{\sqrt{n}} \sum_{i,i',i''=1}^k Y_1^{(i)} Y_1^{(i')} Y_1^{(i'')} \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 \int_0^1 \left\{ \int_{\mathbb{R}^k} \left[\int_{\mathbb{R}^k} \tilde{h}(e^{-s}\sqrt{\frac{n-1}{n}}x + e^{-s}vX_1 + \sqrt{1-e^{-2s}}z) \right. \right. \\
 & \left. \left. (d(Q_{(n-1)}(x) - \Phi(x)) + d\Phi(x)) \right] \cdot D_{ii'i''}\phi(z) dz \right\} dv
 \end{aligned}$$

Again, the inner integral in (3.11) with regard to $Q_{(n-1)} - \Phi$ is estimated by (3.10). Therefore, using (2.3) for the remaining integration with regard to Φ in (3.8), (3.11).

$$\begin{aligned}
 & (3.12) \quad |ET\tilde{h}(S_n)| \leq \frac{1}{\sqrt{n}} \sum_{i,i'=1}^k E|Y_1^{(i')}| \left(\int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 \left[\delta_{n-1} \int_{\mathbb{R}^k} |D_{iii'}\phi(z)| dz + c_0 k e^{2s} (1-e^{-2s}) \right] ds \right) \\
 & + \frac{1}{\sqrt{n}} \sum_{i,i',i''=1}^k E|Y_1^{(i)} Y_1^{(i')} Y_1^{(i'')}| \left(\int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 \left[\delta_{n-1} \left(\int_0^1 (1-v) dv \right) \cdot \int_{\mathbb{R}^k} |D_{ii'i''}\phi(z)| dz + c_0 k e^{2s} (1-e^{-2s}) \right] ds \right).
 \end{aligned}$$

Next, the first two terms on the right in (3.12) may be estimated by using

$$\begin{aligned}
 & (3.13) \quad \int_{\mathbb{R}^k} |D_{ii'i''}\phi(z)| dz = \begin{cases} E|(Z_1^{(i)})^2 - 1| \cdot E|Z_1^{(i')}| \leq 1 & \forall i \neq i', i'' = i \text{ or } i', \\ E|(Z_1^{(i)})^3 - Z^{(i)}| \leq \sqrt{6} & \forall i = i' = i'', \end{cases} \\
 & \int_{\mathbb{R}^k} |D_{ii'i''}\phi(z)| dz = E|Z_1^{(i)} Z_1^{(i')} Z_1^{(i'')}| \leq 1 \text{ if } i, i', i'' \text{ are all distinct.}
 \end{aligned}$$

Finally, note that

$$(3.14) \quad \frac{e^{-s}}{\sqrt{1-e^{-2s}}} \leq \frac{1}{\sqrt{2s}} \quad (s > 0),$$

so that

$$(3.15) \quad \int_0^\infty \frac{e^{-s}}{\sqrt{1-e^{-2s}}} ds = c'_0 < \infty, \quad \int_t^\infty \left(\frac{e^{-s}}{\sqrt{1-e^{-2s}}} \right)^3 ds \leq (2t)^{-\frac{1}{2}}.$$

Hence, using (3.12)-(3.15), together with the estimates

$$E \sum_{i,i',i''=1}^k \left| Y_1^{(i)} Y_1^{(i')} Y_1^{(i'')} \right| \leq k^{\frac{3}{2}} \rho_3, \quad E \sum_{i,i'=1}^k |Y_1^{(i')}| \leq k^{\frac{1}{2}} \rho_3,$$

one has

$$(3.16) \quad |ET_t \tilde{h}(S_n)| \leq c_1 k^{3/2} \rho_3 \left(\frac{\delta_{n-1}}{\sqrt{n}\sqrt{t}} \right) + \frac{c_2 k^{5/2} \rho_3}{n^{1/2}}$$

4. The smoothing inequality and the Estimation of δ_n

Let $\mathcal{H} = \{\mathbf{1}_C, C \in \mathcal{C}\}$, where \mathcal{C} is the class of all Borel convex subsets of \mathbb{R}^k . As before, $\tilde{h} = h - \int h d\Phi$. We also write G_b as the distribution of bW , if W has distribution $G(b > 0)$. Recall that (see 1.15) $T_t \tilde{h}(x) = E \tilde{h}(e^{-t}x + \sqrt{1-e^{-2t}}Z)$, where Z has the standard Normal distribution $\Phi = N(0, \mathbb{I}_k)$, which we take to be independent of S_n . Then

$$(4.1) \quad \begin{aligned} ET_t \tilde{h}(S_n) &= E \tilde{h}(e^{-t}S_n + \sqrt{1-e^{-2t}}Z) = \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \tilde{h}(e^{-t}x + \sqrt{1-e^{-2t}}z) dQ_{(n)}(x) \phi(z) dz \\ &= \int_{\mathbb{R}^k} \tilde{h}d((Q_{(n)})_{e^{-t}} \star \Phi_{\sqrt{1-e^{-2t}}}) = \int_{\mathbb{R}^k} \tilde{h}d((Q_{(n)})_{e^{-t}} - \Phi_{e^{-t}}) \star \Phi_{\sqrt{1-e^{-2t}}} \end{aligned}$$

The introduction of the extra term $\Phi_{e^{-t}} \star \Phi_{\sqrt{1-e^{-2t}}} = \Phi$ does not affect the integration in the last step since $\int_{\mathbb{R}^k} \tilde{h} d\Phi = 0$.

Since the last integration is with respect to the difference between two probability measures, its value is unchanged if we replace \tilde{h} by h . Hence

$$(4.2) \quad ET_t \tilde{h}(S_n) = \int_{\mathbb{R}^k} h d[(Q_{(n)})_{e^{-t}} - \Phi_{e^{-t}}] \star \Phi_{\sqrt{1-e^{-2t}}}.$$

Also the class \mathcal{C} is invariant under multiplication $C \rightarrow bC$ where $b > 0$ is given. Therefore,

$$(4.3) \quad \delta_n = \sup_{h \in \mathcal{H}} |E \tilde{h}(S_n)| = \sup_{h \in \mathcal{H}} \left| \int h d(Q_{(n)} - \Phi) \right| = \sup_{h \in \mathcal{H}} \left| \int h d[(Q_{(n)})_{e^{-t}} - \Phi_{e^{-t}}] \right|.$$

Thus (4.2) is a perturbation (or, smoothing) of the integral in (4.3) by convolution with $\Phi_{\sqrt{1-e^{-2t}}}$. If $\epsilon > 0$ is a constant such that

$$(4.4) \quad \Phi_{\sqrt{1-e^{-2t}}}(\{|z| < \epsilon\}) = \frac{7}{8},$$

then the smoothing inequality below applies, with $\mu = (Q_{(n)})_{e^{-t}}$, $\nu = \Phi_{e^{-t}}$, $K = \Phi_{\sqrt{1-e^{-2t}}}$, $f = h = \mathbf{1}_C$, $\alpha = 7/8$, and ϵ as in (4.4).

Smoothing Inequality

Let μ, ν, K be probability measures on \mathbb{R}^k , $K(\{x : |x| < \epsilon\}) = \alpha > \frac{1}{2}$. Then for every bounded measurable f one has

$$(4.5) \quad \left| \int_{\mathbb{R}^k} f d(\mu - \nu) \right| \leq (2\alpha - 1)^{-1} [\gamma^*(f : \epsilon) + \omega_f^*(2\epsilon : \nu)]$$

where,

$$\begin{aligned} f_\epsilon^+(x) &= \sup\{f(y) : |y - x| < \epsilon\}, \quad f_\epsilon^-(x) = \inf\{f(y) : |y - x| < \epsilon\}, \\ \text{and } \gamma(f : \epsilon) &= \max \left\{ \int_{\mathbb{R}^k} |f_\epsilon^+ d(\mu - \nu)|, \int_{\mathbb{R}^k} |f_\epsilon^- d(\mu - \nu)| \right\}, \\ \gamma^*(f : \epsilon) &= \sup_{y \in \mathbb{R}^k} \gamma(f_y : \epsilon), \quad f_y(x) \equiv f(x + y), \\ \omega_f(x : \epsilon) &= \sup\{|f(y) - f(x)| : |y - x| < \epsilon\}, \quad \omega_f(\epsilon : \nu) = \int \omega_f(x : \epsilon) d\nu(x), \\ \omega^*(f : \epsilon) &= \sup_{y \in \mathbb{R}^k} \omega_{f_y}(\epsilon : \nu). \end{aligned}$$

For a proof of the inequality (4.5) see Bhattacharya and Rao [5], Lemma 11.4. With $h = \mathbf{1}_C$ one gets $h_\epsilon^+ = \mathbf{1}_{C^\epsilon}$, $h_\epsilon^- = \mathbf{1}_{C^{-\epsilon}}$, where $C^\epsilon = \{x : \text{dist}(x, C) < \epsilon\}$, $C^{-\epsilon} = \{x : \text{open ball of radius } \epsilon \text{ and center } x \text{ is contained in } C\}$ are both convex, so that

$$\gamma(h : \epsilon) \leq \max \left\{ \int \mathbf{1}_{C^\epsilon} d[(Q_{(n)})_{e^{-t}} - \Phi_{e^{-t}}] \star \Phi_{\sqrt{1-e^{-2t}}}, \int \mathbf{1}_{C^{-\epsilon}} d[(Q_{(n)})_{e^{-t}} - \Phi_{e^{-t}}] \star \Phi_{\sqrt{1-e^{-2t}}} \right\} \leq \sup_{h \in \mathcal{H}} |ET_t \tilde{h}(S_n)|.$$

Since \mathcal{C} is invariant under translation one then obtains

$$(4.6) \quad \gamma^*(h : \epsilon) \leq \sup_{h \in \mathcal{H}} |ET_t \tilde{h}(S_n)|.$$

Also, letting Z be standard Normal $N(0, \mathbf{1}_k)$,

$$\begin{aligned} \omega_h^*(2\epsilon : \Phi_{e^{-t}}) &= P(e^{-t}Z \in (\partial C)^{2\epsilon}) \\ (4.7) \quad &= P(Z \in e^t(\partial C)^{2\epsilon}) \leq c_3 \sqrt{k} 2\epsilon e^t. \end{aligned}$$

where $c_3 > 0$ is a constant (see Bhattacharya and Rao [5], Theorem 3.1). From (4.4) one gets

$$P\left(\left|\sqrt{1-e^{-2t}}Z\right| < \epsilon\right) = \frac{7}{8}, \quad P\left(|Z| < \frac{\epsilon}{\sqrt{1-e^{-2t}}}\right) = \frac{7}{8}$$

so that $\epsilon/\sqrt{1-e^{-2t}} = a_k$, where a_k satisfies $P(|Z| < a_k) = \frac{7}{8}$. It is simple to check that $a_k = O(\sqrt{k})$, as $k \rightarrow \infty$, and

$$(4.8) \quad a_k \leq c_4 \sqrt{k}, \quad \epsilon = a_k \sqrt{1-e^{-2t}} \leq c_4 \sqrt{k} \sqrt{1-e^{-2t}} \leq c_4 \sqrt{k} \sqrt{2t}$$

Using this estimate of ϵ in (4.9), one obtains

$$(4.9) \quad \omega_h^*(2\epsilon : \Phi_{e^{-t}}) \leq c_5 k \sqrt{t} e^t$$

The smoothing inequality now yields (use (4.3), (4.6), (4.9) in (4.5))

$$(4.10) \quad \delta_n \leq \frac{4}{3} \left[\sup_{h \in \mathcal{H}} |ET_t \tilde{h}(S_n)| + c_5 k \sqrt{t} e^t \right]$$

Now use (3.14) in (4.11) to get

$$(4.11) \quad \delta_n \leq (c_6 k^{3/2} \rho_3) \frac{\delta_{n-1}}{\sqrt{n} \sqrt{t}} + \frac{c_7 k^{5/2} \rho_3}{n^{1/2}} + c_8 k \sqrt{t} e^t.$$

By comparing the first and third terms on the right, an optimal order of t is obtained as

$$t = \min \left\{ 1, \frac{\sqrt{k} \delta_{n-1} \rho_3}{\sqrt{n}} \right\}.$$

It follows that

$$(4.12) \quad \delta_n \leq (c_9 k^{5/4} \rho_3^{1/2}) \frac{\delta_{n-1}^{1/2}}{n^{1/4}} + \frac{c_7 k^{3/2} \rho_3}{n^{1/2}}.$$

Consider now the induction hypothesis : The inequality

$$(4.13) \quad \delta_n \leq \frac{c k^{5/2}}{\sqrt{n}} \rho_3$$

holds for some $n \geq 1$ and an absolute constant $c \geq 1$ specified below. Note that (4.13) clearly holds for $n \leq c^2 k^5 \rho_3^2$. Since $c^2 k^5 \rho_3^2 > k^8$, suppose then (4.13) holds for some $n = n_0 \geq k^8$. Then by (4.12) we can take $n_0 \geq k^3$: under the induction hypothesis, and (4.12),

$$(4.14) \quad \begin{aligned} \delta_{n_0+1} &\leq \frac{c_9 \sqrt{c} k^{\frac{5}{4} + \frac{5}{4}} \rho_3}{((n_0(n_0+1))^{\frac{1}{4}})} + \frac{c_7 k^{3/2} \rho_3}{(n_0+1)^{\frac{1}{2}}} \\ &\leq \frac{c_{10} \sqrt{c} k^{\frac{5}{2}} \rho_3}{(n_0+1)^{\frac{1}{2}}} + \frac{c_7 k^{5/2} \rho_3}{2^9 (n_0+1)^{\frac{1}{2}}} \quad (c_{10} = c_9 + 1, \frac{k^{-1}}{n_0+1} \leq k^{-9} \leq 2^{-9}, \text{ for } k \geq 2). \end{aligned}$$

Now, choose c to be the greater of 1 and the positive solution of $c = c_{10} \sqrt{c} + c_7$, to check that (4.13) holds for $n = n_0 + 1$. Hence (4.13) holds for all n .

We have proved the following result.

Theorem 1 *There exists an absolute constant $c > 0$ such that*

$$(4.15) \quad \delta_n \leq \frac{c k^{\frac{5}{2}} \rho_3}{\sqrt{n}}$$

5. The Non-Identically Distributed Case

For the general case considered in [9], X_j 's ($1 \leq j \leq n$) are independent with zero means and $\sum_{j=1}^n \text{Cov} X_j = \mathbb{I}_k$.

Assume

$$(5.1) \quad \beta_3 \equiv \sum_{1 \leq j \leq n} E \|X_j\|^3 < \infty$$

Let $\{\bar{X}_j : 1 \leq j \leq n\}$ be an independent copy of $\{X_j : 1 \leq j \leq n\}$. Then, writing $S_n = \sum_{j=1}^n X_j$, as before,

$$\begin{aligned}
 E \sum_{i=1}^k D_{ii} \psi_t(S_n) &= E \sum_{j=1}^n \sum_{i,i'=1}^k D_{ii'} \psi_t(S_n) \bar{X}_j^{(i)} \bar{X}_j^{(i')} \\
 (5.2) \quad &= E \left[\sum_{j=1}^n \sum_{i,i'=1}^k D_{ii'} \psi_t(S_n - X_j) \bar{X}_j^{(i)} \bar{X}_j^{(i')} + \sum_{j=1}^n \sum_{i,i',i''=1}^k \bar{X}_j^{(i)} \bar{X}_j^{(i')} \bar{X}_j^{(i'')} \int_0^1 D_{ii'i''} \psi_t(S_n - X_j + v X_j) dv \right],
 \end{aligned}$$

and

$$\begin{aligned}
 (5.3) \quad E[S_n \cdot \nabla \psi_t(S_n)] &= E \left[\sum_{j=1}^n X_j \cdot \nabla \psi_t(S_n) \right] \\
 &= E \left[\sum_{j=1}^n \left\{ X_j \cdot \nabla \psi_t(S_n - X_j) + \sum_{i,i'=1}^k X_j^{(i)} X_j^{(i')} D_{ii'} \psi_t(S_n - X_j) + \sum_{i,i',i''=1}^k X_j^{(i)} X_j^{(i')} X_j^{(i'')} \int_0^1 (1-v) D_{ii'i''} \psi_t(S_n - X_j + v X_j) dv \right\} \right]
 \end{aligned}$$

Subtracting (5.3) from (5.2) and noting that

$$E X_j \cdot \nabla \psi_t(S_n - X_j) = 0,$$

one obtains

$$\begin{aligned}
 (5.4) \quad E T_t \tilde{h}(S_n) &= E \left[\sum_{j=1}^n \sum_{i,i',i''=1}^k \bar{X}_j^{(i)} \bar{X}_j^{(i')} \bar{X}_j^{(i'')} \int_0^1 D_{ii'i''} \psi_t(S_n - X_j + v X_j) dv \right. \\
 &\quad \left. - \sum_{j=1}^n \sum_{i,i',i''=1}^k \bar{X}_j^{(i)} \bar{X}_j^{(i')} \bar{X}_j^{(i'')} \int_0^1 (1-v) D_{ii'i''} \psi_t(S_n - X_j + v X_j) dv \right]
 \end{aligned}$$

The estimation of the conditional expectation of the integrals \int_0^1 in (5.4), given X_j , proceeds as in Section 3 (with X_j in place of X_1). The only significant change is in the normalization in the argument of \tilde{h} (see (3.8) - (3.11)) where, writing N_j as the positive square root of the inverse of $Cov(S_n - X_j)$,

$$\begin{aligned}
 (5.5) \quad E \left[\tilde{h}(e^{-s}(S_n - X_j) + e^{-s}vX_j + \sqrt{1-e^{-2s}z}|X_j) \right] \\
 &= E \left[\tilde{h}(e^{-s}N_j^{-1}(N_j(S_n - X_j)) + e^{-s}vX_j + \sqrt{1-e^{-2s}z}|X_j) \right] \\
 &= \int_{\mathbb{R}^k} \tilde{h}(e^{-s}N_j^{-1}x + e^{-s}vX_j + \sqrt{1-e^{-2s}z}) dQ_{(n-1),j}(x) = \int_{\mathbb{R}^k} \tilde{h}(e^{-s}N_j^{-1}(x + N_j e^s \sqrt{1-e^{-2s}z}) + e^{-s}vX_j) dQ_{(n-1),j}(x),
 \end{aligned}$$

where $Q_{(n)}$ denotes the distribution of $S_n = \sum_1^n X_j$, and $Q_{(n-1),j}$ that of $N_j(S_n - X_j)$, which has mean zero, covariance \mathbb{I}_k . As in Section 3, the last integration is divided into two parts: $d(Q_{(n-1),j} - \Phi)(x) + d\Phi(x)$. Since the class of Borel convex sets is invariant under non-singular affine linear transformations, the integral with regards to $Q_{(n-1),j} - \Phi$ is bounded by δ_{n-1} . For the integral with regards to Φ , we change variables $x \rightarrow y = x + A_j z$, where $A_j = e^{-s} \sqrt{1-e^{-2s}} N_j$. The estimation of the integral now proceeds as in (2.3)–(2.8), with scalar a replaced by the

matrix A_j . The effect of this is simply to change the sum $a^2 \sum_{r,r'} z_r z_{r'} D_{rr'} \phi(y - vaz)$ in (2.4) to

$$\sum_{r,r'=1}^k (A_j z)_r (A_j z)_{r'} D_{rr'} \phi(y - v A_j z)$$

Arguing as in (2.3)–(2.8) one arrives at the upper bound for (5.5) given by

$$c_0'' k \|A_j\|^2 = c_0'' k e^{2s} (1 - e^{-2s}) \|N_j\|^2 \leq c_0'' k e^{2s} (1 - e^{-2s}) (1 - \beta_3^{\frac{2}{3}})^{-1},$$

using

$$\begin{aligned} (5.6) \quad \|N_j\|^2 &= \|(\mathbb{I}_k - \text{Cov} X_j)^{-\frac{1}{2}}\|^2 = \|\mathbb{I}_k - \text{Cov} X_j\|^{-1}, \\ \|\mathbb{I}_k - \text{Cov} X_j\| &= \sup_{|u|=1} u \cdot (\mathbb{I}_k - \text{Cov} X_j) u = \sup_{|u|=1} (1 - E(u \cdot X_j)^2) \\ &\geq 1 - E|X_j|^2 \geq 1 - (E|X_j|^3)^{\frac{2}{3}} \geq 1 - \beta_3^{\frac{2}{3}} \end{aligned}$$

and assuming

$$(5.7) \quad \beta_3 < 1$$

Proceeding as in Section 4 one arrives at the bound:

$$(5.8) \quad \delta_n \leq c k^{\frac{5}{2}} \beta_3.$$

If one takes the absolute constant $c > 1$, then the β_3 may be assumed to be smaller or equal to $c^{-1} k^{-\frac{5}{2}}$, and $(1 - \beta_3^{\frac{2}{3}})^{-1} \leq (1 - \frac{1}{c^{\frac{2}{3}}})^{-1} = c'$. The induction argument is similar.

Remark: If one defines

$$(5.9) \quad \gamma_3 \equiv \sum_{j=1}^n E \left(\sum_{i=1}^k |X_j^{(i)}| \right)^3,$$

then

$$\sum_{j=1}^n \sum_{i,i',i''=1}^k E |X_j^{(i)} X_j^{(i')} X_j^{(i'')}| = \gamma_3,$$

Since γ_3 now replaces $k^{\frac{3}{2}} \beta_3$ in the computations, it follows that

$$(5.10) \quad \delta_n \leq c k \gamma_3$$

Since, $\gamma_3 \leq k^{\frac{3}{2}} \beta_3$, (5.10) provides a better bound than (5.8) or (4.13).

References

- [1] Ball, K. (1993) The reverse isoperimetric problem of Gaussian measure, *Discrete Comput. Geom.*, **10**, pp. 411-420.
- [2] Barbour, A. D. (1988) Stein's Method and Poisson Process Convergence, *Journal of Applied Probability*, Vol. 25, pp. 175-184
- [3] Bentkus, V. (2003) On the dependence of the Berry-Esseen bound on dimension. *J. Stat. Plan. Inf.*, **113**, pp. 385-402.
- [4] Bhattacharya, R. N. (1982) On the functional central limit theorem and the law of the iterated logarithm for Markov processes, *Zeit. Wahr. Ver. Geb.* 60, 185-201.
- [5] Bhattacharya, R. N. and Ranga Rao, R. (1976) *Normal Approximation and Asymptotic Expansions*, Wiley, New York.
- [6] Bhattacharya, R.N. and Waymire, E.C. (2009) *Stochastic Processes with Applications*. SIAM, Philadelphia.
- [7] Diaconis, P. and Holmes, S. (2004) *Stein's Method: Expository Lectures and Applications*, IMS Lecture Notes, **46**, pp. 45-68.
- [8] Holmes, S. (2004) Stein's method for birth and death chains. in *Stein's Method: Expository Lectures and Applications*, IMS Lecture Notes, **46**, pp. 45-68.
- [9] Götze, F. (1991) On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724-739.
- [10] Raic, M (2004) On the rate of convergence of the multivariate CLT, Personal Communication.
- [11] Rinott, Y. and Rotar, V. (1997). On coupling constructions and rates in the clt for dependent summands with applications to the anti-voter model and weighted U-statistics. *Ann. Applied Probability* **7**, 1080-1105.
- [12] C. Stein (1986). *Approximate Computation of Expectations*. IMS, Hayward, California.